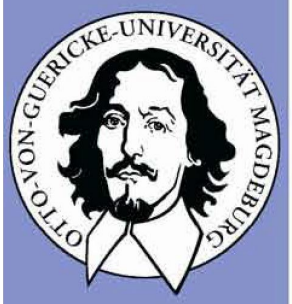


Nr.: FIN-02-2016

Discriminative Star Coordinates

Yunhai Wang, Feiping Nie, Dirk J. Lehmann, Minglun Gong

Visual Computing

Technical report

Fakultät für Informatik
Otto-von-Guericke-Universität Magdeburg

Nr.: FIN-02-2016

Discriminative Star Coordinates

Yunhai Wang, Feiping Nie, Dirk J. Lehmann, Minglun Gong

Visual Computing

Technical report (Internet)
Elektronische Zeitschriftenreihe
der Fakultät für Informatik
der Otto-von-Guericke-Universität Magdeburg
ISSN 1869-5078

Fakultät für Informatik
Otto-von-Guericke-Universität Magdeburg

Impressum (§ 5 TMG)

Herausgeber:
Otto-von-Guericke-Universität Magdeburg
Fakultät für Informatik
Der Dekan

Verantwortlich für diese Ausgabe:
Otto-von-Guericke-Universität Magdeburg
Fakultät für Informatik
Dirk J. Lehmann
Postfach 4120
39016 Magdeburg
E-Mail: dirk@isg.cs.uni-magdeburg,de

http://www.cs.uni-magdeburg.de/Technical_reports.html

Technical report (Internet)
ISSN 1869-5078

Redaktionsschluss: 13.06.2016

Bezug: Otto-von-Guericke-Universität Magdeburg
Fakultät für Informatik
Dekanat

Discriminative Star Coordinates

Yunhai Wang, Feiping Nie, Dirk J. Lehmann, and Minglun Gong

Abstract—We propose discriminative star coordinates (DSC) in order to preserve class structures of high-dimensional data within the related 2D projection. Our novel visualization approach for high-dimensional data utilizes the subspace selected by a linear discriminant analysis (LDA) to visualize data-specific class separation. While LDA can handle labeled data, we introduce generalized unsupervised LDA to facilitate dealing even with arbitrary kinds of unlabeled data. For this, data clustering and subspace selection are coherently combined. Since LDA projects k classes of m-dimensional data into $k-1$-dimensional space (or less), our concept can be extended to a discriminative star coordinates matrix that allows us to explore the majority of important cluster structures of the data at once. To support the users' interactive exploration of classes of interest, we present a two-stage exploration scheme where DSC and principle component star coordinates (PCSC) configured by principle component analysis (PCA) are integrated together. Meanwhile, a set of structure-aware interactions are provided. This includes interactively steering the number of relevant clusters, structure-aware axes manipulation, and a linked parallel coordinates view. We demonstrate the effectiveness of discriminative star coordinates in visual cluster analysis with a couple of experiments for synthetic and real high-dimensional data.

Index Terms—Star Coordinates, Multivariate Visualization, Linear Discriminant Analysis

1 Introduction

Technology advances let high-dimensional data frequently occur in many application domains, such as information retrieval, computational biology, and text mining. To help the user to gain insights from such data, many techniques have been proposed. Among them, visual cluster analysis is an effective way to facilitate exploratory analysis [6], which tightly combines cluster analysis techniques and interactive visualization methods. However, clustering such high-dimensional data is a big challenge due to the curse of dimensionality [30]. This restricts the interactive visual cluster analysis of high-dimensional data.

A common approach to address this problem is to use unsupervised dimension reduction techniques, to project the high-dimensional data onto a low-dimensional subspace before visual cluster analysis. Once a suitable low dimensional (i.e., 2D or 3D) subspace is obtained, many existing visual cluster analysis techniques [30, 31] can be applied. However, one subspace determined by dimension reduction does not include all cluster structures of high-dimensional data. Moreover, some widely used dimension reduction methods in the visualization community, such as principal component analysis (PCA) [17] and multidimensional scaling (MDS) [21], treat each dimension uniformly. Unfortunately, a large number of dimensions in high-dimensional data are irrelevant [39].

Another approach is to directly map m-dimensional variable space to a 2-dimensional plane by using visualization schemes, such as scatterplot matrix [2], parallel coordinates [13], and star coordinates [18, 19]. Although scatterplot matrix and parallel coordinates both have their own advantages, they also treat each dimension uniformly and thus they are generally not effective in discovering cluster structures of high-dimensional data [6]. Moreover, they are limited by the dimensionality of the data (usually 20 dimensions at most).

Although star coordinates are a variant of circular parallel coordinates, they allow the user to assign a 2D weight vector to each dimension so that irrelevant dimensions can be suppressed. Based on a $2 \times m$ weight/proejction matrix, they map m-dimensional data onto a 2D linear subspace. In star coordinates, each dimension corresponds to an axis which is arranged on a circle with the origin at the center. By allowing the user to interactively adjust each axis, star coordinates can be taken as a form of interactive dimension reduction [9]. With carefully refined projection matrices, the cluster structures of high-dimensional data can be revealed. However, designing such projection matrices is a tedious trial-and-error process. Moreover, even given such projection matrices, it is hard for the user to explore the class structures in star coordinates without cluster analysis, especially when data is large.

We propose discriminative star coordinates (DSC), which enable a semiautomatic visual cluster analysis of high-dimensional data. First, we show that the projection of star coordinates is a special case of linear dimension reduction and then we configure the star coordinates with the subspace extracted by linear discriminant analysis (LDA) [16]. LDA is a well-developed linear dimension reduction method, which is able to find the best subspace to discriminate different clusters. However, LDA is a supervised dimension reduction method and can only work for labeled data [7]. To deal with general unlabeled data, we introduce unsupervised LDA (ULDA) [11], which adaptively selects the most discriminative subspace by combining LDA and k-mean clustering. Since it uses discriminative subspace for clustering, irrelevant dimensions can be depressed. For the data with k classes, LDA or ULDA provides $k-1$ dimensions subspace. When $k > 3$, 2D star coordinates cannot show all of them. We propose a discriminative star coordinate matrix, which consists of all possible discriminative star coordinates of $k-1$ dimension subspace.

This automatic subspace selection and axes configuration scheme provides an overview of the class structure in the data to the user. To further reveal class structures and their relationship to dimensions, we present a two-stage exploration scheme, where DSC is coherently combined with principle component star coordinates (PCSC) configured by PCA. In this scheme, DSC is first used to explore the separation between different classes and then PCSC is applied to reveal the dominant dimensions of classes. Under this scheme, several structure-aware interactions are provided. First, we allow the user to adjust the number of clusters and compare different star coordinates with morphing. Second, the user can interactively manipulate each axis to explore the discriminative ability of each dimension and the correlation between different dimensions. Last, a linked parallel coordinates view is provided to help the user explore the class distribution at the original dimensions. In summary, the main contributions of this paper include:

- We build the relationship between star coordinates and linear dimension reduction and propose discriminative star coordinates

- *Yunhai Wang is with the Interdisciplinary Research Center (IRC), Shandong, China*
- *Feiping Nie is with Department of Computer Science and Engineering, University of Texas, Arlington.*
- *Dirk J. Lehmann is with the Department of Simulation and Graphics, University Magdeburg, Germany E-mail: dirk@isg.cs.uni-magdeburg.de.*
- *Minglun Gong is with the Memorial University of Newfoundland*

which can reveal the cluster structures of high-dimensional data in 2D space;

- We introduce unsupervised LDA to automatically select the most discriminative subspace while identifying meaningful clusters from the high-dimensional data;
- We present a two-stage structure exploration scheme to help the user explore the relation between the cluster structures and the dimensions of high-dimensional data.

The rest of the paper is organized as follows. We provide a brief summary of the related work in Section 2. The LDA and ULDA supported discriminative star coordinates are described in Section 3. The structure-aware interactions are introduced in Section 4. After presenting our case studies and discussions in Section 5 and Section 6, we conclude the paper in Section 7.

2 Related work

In this section, we review related multidimensional data visualizations in cluster analysis. A subset of those techniques - namely the family of star coordinate approaches - is subsequently considered. We close the discussion by techniques on subspace clustering analysis.

2.1 Multidimensional Data Visualization

Multidimensional data visualization - also known as multivariate projections - are techniques which project high-dimensional data from the data space onto a lower-dimensional (usually) 2D visualization space. They facilitate insights into the data but generally cause a loss of information. To overcome this, they often come with interaction techniques making a visual search for relevant subspaces feasible.

By showing all pairwise combinations of scatterplots, a scatterplot matrix [2] provides to reveal all pairwise correlations. Parallel coordinates [13] represent the dimensions as a set of parallel axes and render each data tuple as a polyline. These methods are tailored to visualize correlations and trends, but are not effective for cluster analysis. Recently, researchers tried to enhance the cluster analysis of these methods [15,41]. However, they can only handle 20 dimensions at most due to their poor scalability, except [22]. A complete review of these methods is beyond the scope of this paper; please see Keim et al. [1] for more details.

Dimension reduction is another widely used method to visualize data with large dimensions. It is achieved by first projecting m-dimensional points to 2D points with unsupervised dimension reduction methods and then visualizing these points with 2D scatterplots. The most commonly used dimension reduction methods for visual data analysis include PCA [17], LDA [16], and many variants of MDS [4]. PCA is an unsupervised method that purses a subspace preserving the maximal data variances, while LDA selects the best subspace to separate different classes from a labeled data set. To combine the advantages of these two methods, Choo et al. [7,8] propose a two-stage framework for the visualization of labeled data. They first use LDA to obtain reduced dimensional data, which preserves the cluster structure in the data, and then map the data to a 2D scatterplot with PCA. Oesterling et al. [29] use a similar two-stage framework to visualize classified document collections. Unlike these two methods, MDS takes a matrix of pair-wise distances between all data pairs and computes a position for each point in low-dimensional space where the distances between data pairs are preserved. By converting the data into low dimensions, dimension reduction provides a means to explore the structures hidden in the data. However, the original dimensionality information is lost and thus the result is hard to explain. Value and relation (VaR) display [38] proposed by Yang et al. is an exception that visualizes the dimension correlation by mapping dimensions to a 2D space with MDS. However, it does not support cluster analysis.

The same problem in scatterplot matrix, parallel coordinates and dimension reduction methods is that they all treat each dimension uniformly. However, a large number of dimensions are irrelevant in high-dimensional data [39]. LDA is an exception, but it can only work on the labeled data [7, 10]. In this paper, we introduce unsupervised LDA to handle unlabeled data.

2.2 Star Coordinates

The method of star coordinates is proposed by Kandogan et al. [18,19]. They are defined by uniformly arranging m coordinate axes on a circle with the origin at the center. Traditionally, star coordinates are viewed as a variant of parallel coordinates. However, they do not show the exact value of dimensions and instead represent a 2D linear embedding of the original data by using a projection matrix defined by m axes. In this paper, we explain how this representation is related with linear dimension reduction.

Star coordinates have been used for visual classification [34] and volume data exploration [3]. Recently, they have been extended in various ways. Coorprider and Burton [9] extend star coordinates into three dimensions, and Shaik and Yeasin propose to automatically find the best configuration of 3D star coordinates based on MDS results [32]. They are also extended to explore continuous attribute spaces [27] and orthographic projection [23]. By interpreting the projection of star coordinates as a subspace, we extend the concept of star coordinates to discriminative projections.

2.3 Subspace Clustering

Cluster analysis of high-dimensional data suffers from two problems: the existence of irrelevant dimensions and the curse of dimensionality [39]. The irrelevant dimensions can confuse clustering algorithms by masking clusters in noisy data, while the curse of dimensionality makes the distance measure become increasingly meaningless. Due to them, the traditional full-dimensional clustering algorithms become impractical for the analysis of high-dimensional data. Although feature selection methods have been proven to be somewhat effective in improving cluster quality, they are limited in discovering clusters that exist in multiple, overlapping subspaces [20]. Recent research has suggested the subspace cluster analysis to overcome the inherent problem in traditional clustering algorithms and feature selection methods.

Subspace clustering aims to discover the clusters embedded in multiple, overlapping subspaces of high-dimensional data. The early subspace clustering algorithms focus on selecting axis parallel subspaces which consist of a small number of original dimensions [14]. However, this kind of subspace does not have enough flexibility to handle clusters which extend along a mixture of directions. To find such arbitrarily oriented clusters, many algorithms have been proposed to discover arbitrarily oriented subspaces [35]. One of the representative methods is LDA, which seeks a subspace where all clusters are well-separated. To adapt this supervised subspace learning method for general unlabeled data, Ding and Li propose unsupervised LDA (ULDA) [11]. By combining linear LDA and K-means clustering in a coherent framework, it can simultaneously select the subspace and cluster the data. This provides the initial configuration of our proposed discriminative star coordinates.

To support subspace exploration for high-dimensional data, Tatu et al. [33] propose a semi-automatic approach which first searches all possible subspaces for a given data set, and then applies a visual analysis method to explore the obtained subspaces. By organizing the subspaces as a tree, Yuan et al. [40] present a dimension projection matrix/tree that enables the user to understand the relationship between different subspaces. These two methods both separate the subspace selection and data clustering into two different stages. For large data, it is quite hard for the user to quickly discover all interesting structures from all possible subspaces. In contrast, our discriminative star coordinates start from the result of discriminative analysis, where subspace searching and data clustering are integrated into one coherent framework.

3 Discriminative Star Coordinates

Given a set of data records $\mathbf{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$, $\mathbf{x}_i \in \mathbb{R}^m$, which has been centered and normalized in the preprocessing step, star coordinates

project each mD point $\mathbf{x}_i$ to a 2D point $\mathbf{x}'_i$ with a matrix multiplication

$$\mathbf{x}'_i = \mathbf{G}^T \cdot \mathbf{x}_i, \tag{1}$$

where $\mathbf{G}^T = (\mathbf{g}_1, \cdots, \mathbf{g}_m)$, $\mathbf{g}_j \in \mathbb{R}^2$ is a set of 2D vectors of weights. The vector $\mathbf{g}_j$ is not only the 2D weight of the j^{th} dimension of $\mathbf{x}_i$, but also the j^{th} axis within star coordinates where the origin is the zero vector. Initially, $\mathbf{G}$ can be automatically set with equally radial alignment [18], here referred to as standard configuration, and then it can be changed by interactively moving $\mathbf{g}_j$.

By generalizing the 2D projection space to $l (l < m)$ dimensional space, linear dimension reduction takes the same method to project each mD point data record $\mathbf{x}_i$ to lD $\mathbf{x}'_i$. Rather than interactively finding the proper $\mathbf{G}$ in star coordinates, linear dimension reduction aims to automatically find an optimal $\mathbf{G}$, which can almost preserve the structure of relevant patterns and clusters of high-dimensional data in a low-dimensional space [37]. Thus, linear dimension reduction provides an appropriate projection matrix to initialize star coordinates. We denote the projection matrix of star coordinates by $\mathbf{G}_{sc}$.

As one of the linear dimension reduction methods, linear discriminative analysis (LDA) [12] aims to find a class-preserving projection matrix $\mathbf{G}$ which can separate different classes of the data in a low-dimensional space. Hence, using this approach to configure star coordinates yields discriminative star coordinates (DSC) where different classes are well-separated in star coordinates. How this approach works is briefly explained subsequently.

3.1 Linear Discriminant Analysis

Assume $\mathbf{X}$ consists of a number of k classes, the corresponding label is $\mathbf{y} = \{y_1, \cdots, y_n\}$, where $y_i \in \{1, \cdots, k\}$. In LDA, three scatter matrices are defined, namely the total scatter $\mathbf{S}_t$, the between-cluster scatter $\mathbf{S}_b$, and the with-in scatter $\mathbf{S}_w$, as follows [12]:

$$\mathbf{S}_t = \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T \tag{2}$$

$$\mathbf{S}_b = \sum_{i=1}^{k} \frac{n_i}{n} (\mu_i - \mu)(\mu_i - \mu)^T \tag{3}$$

$$\mathbf{S}_w = \sum_{i=1}^{k} \sum_{y_j = i} (\mathbf{x}_j - \mu_i)(\mathbf{x}_j - \mu_i)^T \tag{4}$$

where $\mathbf{x}_i \in \mathbb{R}^m$, $\mu = \sum_{i=1}^{n} \mathbf{x}_i / n$ is the global mean of the whole data records, n_i is the number of the records of the i^{th} class, and $\mu_i = \sum_{y_j = i} \mathbf{x}_j / n_i$ is the mean of the i^{th} class. It can be easily derived that $\mathbf{S}_t = \mathbf{S}_w + \mathbf{S}_b$. The within-cluster scatter of the projected $\mathbf{X}$ can be expressed as:

$$\begin{aligned} \mathbf{S}'_w &= \sum_{i=1}^{k} \frac{n_i}{n} (\mathbf{x}'_i - \mu'_{y_i})(\mathbf{x}'_i - \mu'_{y_i})^T \\ &= \sum_{i=1}^{k} \mathbf{G}^T (\mathbf{x}_i - \mu_{y_i})(\mathbf{x}_i - \mu_{y_i})^T \mathbf{G} \\ &= \mathbf{G}^T S_w \mathbf{G}, \end{aligned}$$

which gives an implicit description for the projection matrix $\mathbf{G}$. Similarly, $\mathbf{S}'_b = \mathbf{G}^T \mathbf{S}_b \mathbf{G}$.

To characterize classes in a low-dimensional space, LDA attempts to find $\mathbf{G}$ that can maximize the between-class scatter $\mathbf{S}'_b$ and minimize the within-class scatter $\mathbf{S}'_w$. Hence, an optimal transformation $\mathbf{G}$ would maximize $\text{trace}(\mathbf{S}'_b)$ and minimize $\text{trace}(\mathbf{S}'_w)$:

$$\max_{\mathbf{G}} \frac{\text{trace}(\mathbf{S}'_b)}{\text{trace}(\mathbf{S}'_w)} = \frac{\text{trace}(\mathbf{G}^T \mathbf{S}_b \mathbf{G})}{\text{trace}(\mathbf{G}^T \mathbf{S}_w \mathbf{G})} \tag{5}$$

However, this problem does not have a closed-form global optimal. Often, it is approximated by [36]

$$\max_{\mathbf{G}} \text{trace} \frac{\mathbf{G}^T \mathbf{S}_b \mathbf{G}}{\mathbf{G}^T \mathbf{S}_w \mathbf{G}}, \tag{6}$$

where the columns of $\mathbf{G}$ are the eigenvectors of the non-zero eigenvalues of the following generalized eigenvalue problem

$$\mathbf{S}_b \mathbf{g} = \lambda \mathbf{S}_w \mathbf{g}. \tag{7}$$

If the data consists of k classes, there are $k-1$ non-zero eigenvalues, i.e., $\mathbf{G}$ consists of $k-1$ eigenvectors. When $k = 3$, $\mathbf{G}$ is a $2 \times m$ matrix and can be directly used to configure star coordinates. Figure 1 shows two examples with 3 classes.

3.2 Configuring 2D Discriminative Star Coordinates

To consistently encode class structures and LDA-extracted subspace into star coordinates, two visual encoding methods have been proposed. First, we compute the convex hull of the points from each class and smooth its contour. The contour is colorized with the color of its corresponding class. Second, we select any pair of eigenvectors from $\mathbf{G}$ to construct a $2 \times m$ matrix $\mathbf{G}_{sc}$, and each column of this matrix is used to set the position of the corresponding axis. By default, we use the eigenvectors of the first two leading eigenvalues to set the projection matrix $\mathbf{G}_{sc}$.

Figure 1 shows the comparison of star coordinates configured by standard configuration [18], PCA [23], and LDA for the *Iris* data set with $m = 4$ and the *Wine* data set with $m = 13$, respectively. These data sets both consist of three classes shown in different colors. It can be seen that LDA-configured DSC clearly well separates 3 classes, while there are overlappings between classes in the other two star coordinates.

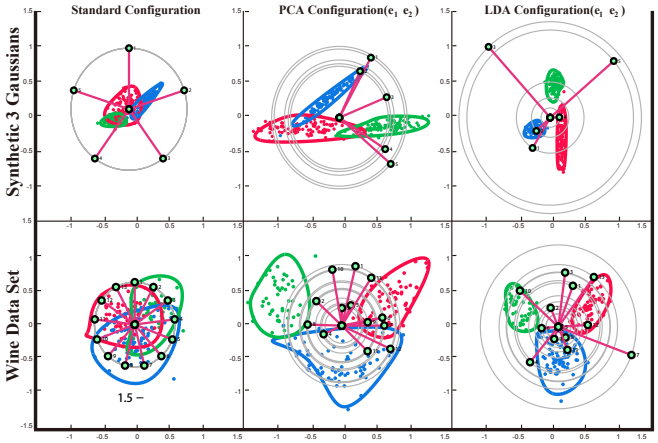


Fig. 1. (left) Standard Configuration, (middle) PCA configuration to the first two largest eigenvalues, and (right) DSC configured by LDA obtained eigenvector of the first two largest eigenvalues, for two data sets: (top) the Iris data set with m=4 dimensions and (bottom) the Wine data set with m=13 dimensions. Both data sets consist of 3 classes, with each class being color-coded. It can be seen that the LDA configuration can best discriminate clusters in both tests.

3.3 Unsupervised Linear Discriminant Analysis

LDA is a supervised dimension reduction method, which requires that the data has class labels. However, most of the data is unlabeled in real-world applications [5] and thus DSC cannot only be applied to visualize general data. A straightforward way to address this problem is to perform K-means clustering [25] in high-dimensional data space and then use the obtained labels for LDA. However, direct clustering of high-dimensional data cannot accurately identify the cluster structure, because many irrelevant dimensions in high-dimensional data may confuse clustering algorithms [30]. Accordingly, we introduce unsupervised LDA (ULDA) [11], which jointly performs K-means clustering and LDA.

Our main goal is to select proper $\mathbf{G}$ in an unsupervised way so that the cluster structures can be revealed in a low-dimension space.

Hence, the optimization function in ULDA is the same with LDA

$$\max_{\mathbf{G},\mathbf{y}} \frac{\text{trace}(\mathbf{G}^T\mathbf{S}_b\mathbf{G})}{\text{trace}(\mathbf{G}^T\mathbf{S}_w\mathbf{G})}, \tag{8}$$

However, this optimization involves two sets of unknown variables: $\mathbf{G}$ and $\mathbf{y}$ and is generally hard to solve. We design a two-step approach to approximate the optimization by alternatively fixing $\mathbf{G}$ and $\mathbf{y}$. When $\mathbf{y}$ is fixed, $\mathbf{G}$ can be found by a standard LDA procedure, as shown in Equation 7. When $\mathbf{G}$ is fixed, Equation 8 turns out to be

$$\begin{aligned}\max_{\mathbf{y}} \frac{\text{trace}(\mathbf{G}^T\mathbf{S}_b\mathbf{G})}{\text{trace}(\mathbf{G}^T\mathbf{S}_w\mathbf{G})} &= \frac{\text{trace}(\mathbf{G}^T(\mathbf{S}_t-\mathbf{S}_w)\mathbf{G})}{\text{trace}(\mathbf{G}^T\mathbf{S}_w\mathbf{G})} \\ &= \frac{\text{trace}(\mathbf{G}^T\mathbf{S}_t\mathbf{G})}{\text{trace}(\mathbf{G}^T\mathbf{S}_w\mathbf{G})} - 1 \end{aligned} \tag{9}$$

Since $\text{trace}(\mathbf{G}^T\mathbf{S}_t\mathbf{G})$ is a constant, it becomes the minimization of $\text{trace}(\mathbf{G}^T\mathbf{S}_w\mathbf{G})$:

$$\min_{\mathbf{y}} \text{trace}(\mathbf{G}^T\mathbf{S}_w\mathbf{G}) = \sum_{i=1}^{k}\sum_{y_j=i} \mathbf{G}^T||\mathbf{x}_j-\mu_i||^2\mathbf{G} \tag{10}$$

This is equivalent to perform K-means clustering on the space of $\mathbf{G}^T\mathbf{X}$. Thus, when $\mathbf{G}$ is fixed, $\mathbf{y}$ can be obtained by performing K-means on the projected low-dimensional space $\mathbf{G}^T\mathbf{X}$. Initially, $\mathbf{G}$ is constructed by PCA. Since K-means is sensitive to the initial centers, we run K-means multiple times with randomly selected centers and then choose the one which has the smallest within-cluster variation. When $\mathbf{y}$ is fixed, $\mathbf{G}$ can be solved with the standard LDA method (Equation 7).

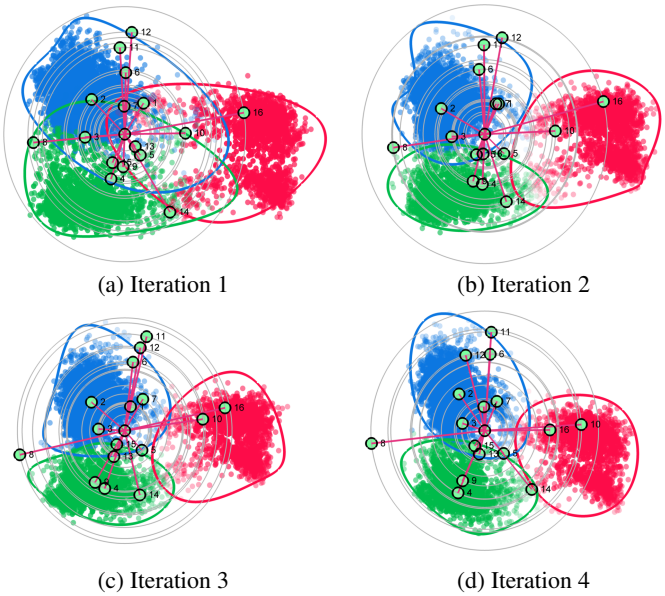
In general, unsupervised LDA can be solved with a two-step iterative algorithm. Specifically, starting with $\mathbf{G}$ initialized by PCA, we alternate between finding $\mathbf{y}$ with K-means clustering and searching $\mathbf{G}$ by standard LDA procedure. Hence, the time complexity of unsupervised LDA is $O(mnt)$ for K-means clustering and $O(p^2nt)$ for LDA computation, where m, n and t are the numbers of dimensions, points and iterations, respectively. In our experiment, we find out that this algorithm can achieve convergence in less than 10 iterations. Figure 2 displays the DSCs of four iterations of visualizing the *digits* data set with $k=3$. We can see that the separation of cluster structures is gradually improved as the number of iteration increases. Within four iterations, these three clusters are well separated in star coordinates.

As configuring 2D DSC with LDA, we use the eigenvectors obtained from Equation 7 to initialize star coordinates. By default, the star coordinates are configured by the eigenvectors of the first two leading eigenvalues. Unlike labeled data visualization, we also visualize the membership of each point belonging to its corresponding cluster by setting its opacity based on its distance to its corresponding cluster center. In Figure 2, each point is visualized in this way.

3.4 Discriminative Star Coordinates Matrix

As pointed out by [7], the eigenvectors of the first two leading eigenvalues of $\mathbf{G}$ cannot reveal all class information when $k>3$. To resolve this issue, one approach is to allow the user to choose any pair of eigenvectors from $\mathbf{G}$ while providing morphing [23] between star coordinates with different eigenvectors configurations. However, this method does not give an overview of class structures. Another method is to use multiple DSCs where each DSC shows a local view of three classes [10]. By linking these DSCs, the user can get an idea of how one class interacts with another class, but it is hard to get the complete spatial relationship among multiple classes.

Inspired by the dimension projection matrix [40], we propose the discriminative star coordinates matrix (DSCM), where each pair of eigenvectors is chosen to configure different star coordinates. Compared to the dimension projection matrix [40], DSCM not only gives a complete overview of the class structures, but also reveals how the cluster relates to dimensions. For more detail, see Section 4.2. When k is very large, we only show the DSCM constructed by the eigenvectors of the first five leading eigenvalues, because the discriminative ability of the last eigenvectors is small.



(a) Iteration 1 (b) Iteration 2

(c) Iteration 3 (d) Iteration 4

Fig. 2. The star coordinates of four iterations of using unsupervised LDA to visualize the *digits* data set with $k=3$. (a)The star coordinates of the first iteration where three classes are mixed together. (b) The star coordinates of the second iteration, where three classes become compact although they still overlap each other. (c,d) The star coordinates of the third and fourth iterations where three classes are gradually separated.

Figure 3 shows an example for the synthetic *four-Gaussian* with $m=4$. The star coordinates configured by the first two leading eigenvectors e_1 and e_2 separate the data into three clusters, which correspond to the green class, the magenta class and the mix of red and blue classes, respectively. The weights of axes indicate that these three separated classes have large differences on the axes 1, 3 and 4. The star coordinates configured by the eigenvectors e_1 and e_3 can better separate red and blue classes, but do not discriminate green and magenta classes. Thus, we can conclude that the red and blue classes are different in axes 1 and 3. It is worth noting that the length of the second axis is very small in all star coordinates, suggesting that the second axis is not useful for visual cluster analysis. This is consistent with the mechanism used in our data generation.

4 Two-Stage Structure Exploration

In the following, we propose a framework for an iterative structure exploration process, i.e., a pure visual analytics tool for intuitive and effective visual cluster detection. With a class-preserving projection, DSC can preserve the inter-class structure present in the original high-dimensional data. However, it does not characterize the structure of the data. On the other hand, the star coordinates configured by PCA [23] can preserve the shape of the data in the low-dimensional space, though it cannot characterize different classes. We name the star coordinates with PCA configuration as principal component star coordinates (PCSC). By default, PCSC is configured by the eigenvectors of the first two leading eigenvalues of the generalized eigenvalue problem:

$$\mathbf{S}_t\mathbf{g} = \lambda\mathbf{g}. \tag{11}$$

where $\mathbf{S}_t$ is the covariance of the data defined in Equation 2. Figure 4(d) illustrates an example of PCSC, where three ellipses reveal the major shape variations of three ellipsoids shown in Figure 4(a).

In order to reveal more structures of the data, we propose a two-stage structure exploration scheme by integrating LDA/ULDA-supported DSC and PCA-supported PCSC together.

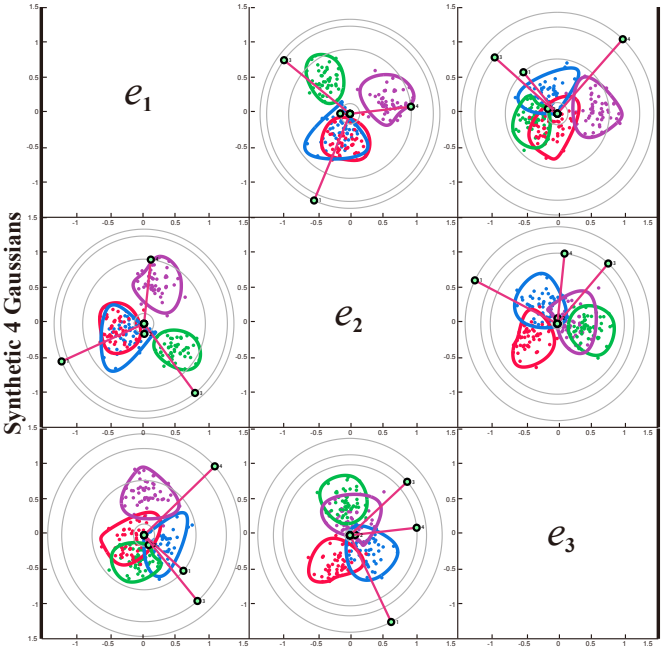


Fig. 3. A DSCM of the synthetic *four-Gaussian* data set with $m = 4$. Note that the star coordinates at location m_{ij} of the matrix are a reflection of the star coordinates at location m_{ji}.

In this scheme, (i) the user can easily get an overview of separated classes from DSC and then (ii) explore the structures of data and each class with PCSC views.

Figure 4 shows an example of exploring a synthetic *three-Gaussian* data set with $m = 3$ (Figure 4(a)). Starting from the DSC view (Figure 4(c)), the user learns that the data form three separated clusters. However, in an effort to separate data from different classes, the DSC uses a very short axis for the Y-dimension since it has a much lower discriminative ability than the X and Z dimensions. As a result, the DSC view is similar to what we would get by projecting the data onto the X-Z plane. While this is the optimal strategy for separating these three input ellipsoids, it also hides the structures of data along the Y-dimension, i.e., the data sampled from the red ellipsoid are rather projected onto a circle than an ellipse. As a complement to DSC, the PCSC view (Figure 4(d)) shows three elongated ellipses, which are consistent with the shapes of the ellipsoids in Figure 4(a). Moreover, we can also observe that the axis of the Y-dimension is the longest among the three, which indicates that the data have the largest variance along the Y-dimension. To allow users to explore the structure of the data from each class, we further generate the PCSC views for individual classes, see Figure 4(e,f,g). From these views, the user can clearly see the elliptical shape of data in red and green classes. Nevertheless, the records in the blue class are projected into a circle in Figure 4(f), even though they are sampled from a nearly flat ellipsoid. This is because the PCSC uses the two dominant directions for projection and hence hides the structure of the data along the direction with the smallest eigenvalue, while combining Figure 4(c) and Figure 4(f) gives us a complete shape structure of the blue ellipsoid.

Under this two-stage exploration scheme, the user can interactively explore the structures of the high-dimensional data from three aspects: class structures, the relationship between class structures and data dimensions, and the distribution of classes in the original data. These three goals are achieved by steering the number of clusters in DSC, manipulating axes in DSC and PCSC, and visualizing the selected classes and dimensions with linked parallel coordinates.

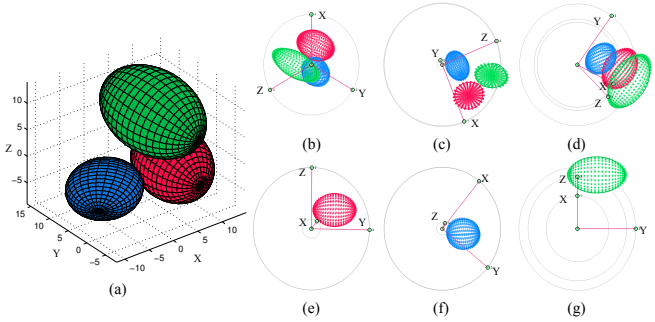


Fig. 4. Exploration of the synthetic *three-Gaussian* data set using the two-stage scheme. (a) The input data are sampled from three adjacent but separated Gaussian ellipsoids in 3D space. (b) The star coordinates with standard configuration mix the three ellipses together. (c) Three separated but adjacent ellipses in 2D DSC, where the axis length of the second dimension is the smaller one compared to the other two dimensions. (d) Three elongated ellipses overlap in 2D PCSC, where the axes lengths of three dimensions are close. (e,f,g) The 2D PCSCs of three different Gaussians, where the second dimension is the significant dimension of all Gaussians.

4.1 Steering Number of Clusters

Given unlabeled data, the user often knows little or nothing about structures underlying the data. Without any prior knowledge, it is hard to determine the number of clusters to be used for generating DSC views. To address this problem, we take an interactive approach to explore the proper number of clusters and to update the DSC and PCSC views accordingly. To help the user intuitively compare the changes of the class structure, we implement the orthography-preserving morphing [23] approach for interpolating between two star coordinates with different configurations. It is particularly useful in learning how the degree of separation between clusters is changed after setting a different number of clusters.

Figure 5 illustrates an example of exploring the *satimage* data set with two different numbers of clusters $k = \{3,4\}$. Figure 5 (a) shows three separated classes and Figure 5 (d,e) reveals four classes in DSC views, which are configured by eigenvectors (e_1, e_2) and (e_1, e_3), respectively. Comparing Figure 5 (b) with Figure 5 (d), the user may conclude that the magenta cluster is generated by splitting the blue cluster in Figure 5 (a). On the other hand, the magenta and red clusters are close to each other in Figure 5 (e), making it hard to draw a conclusion. The PCSC views for 3 and 4 clusters shown in Figure 5 (c and f) provide additional information about the data structure. They show that the majority of the data in the magenta cluster come from the red cluster in the 3-cluster result, with the rest from both blue and green clusters.

4.2 Axes Manipulation

In DSC, the axis length represents the discriminative ability of the corresponding dimension, whereas the axis length of PCSC indicates the variance of the data along the corresponding dimension. This difference is clearly illustrated by the axis length of the Y-dimension in Figure 4(b,c). The axes in the PCSC view also reveal the correlations between different dimensions. If two axes have similar lengths, the angle between them reflects their correlation. If the angle is small, they are positively correlated and thus one of them may be redundant; if the angle is around 90°, they are not correlated; otherwise, they are negatively correlated. According to this guideline, we can see that the first and third dimensions are highly correlated (Figure 4(f)). This is consistent with the green ellipsoid shown in Figure 4(a), whose radii along the first and third dimension are quite close.

To further help the user understand such relationship, scaling and rotation of axes are provided [18]. Figure 6 illustrates the axis scaling and rotation operations on the DSC view (Figure 6(a)) of the *satimage* data set configured with the number of clusters $k = 5$

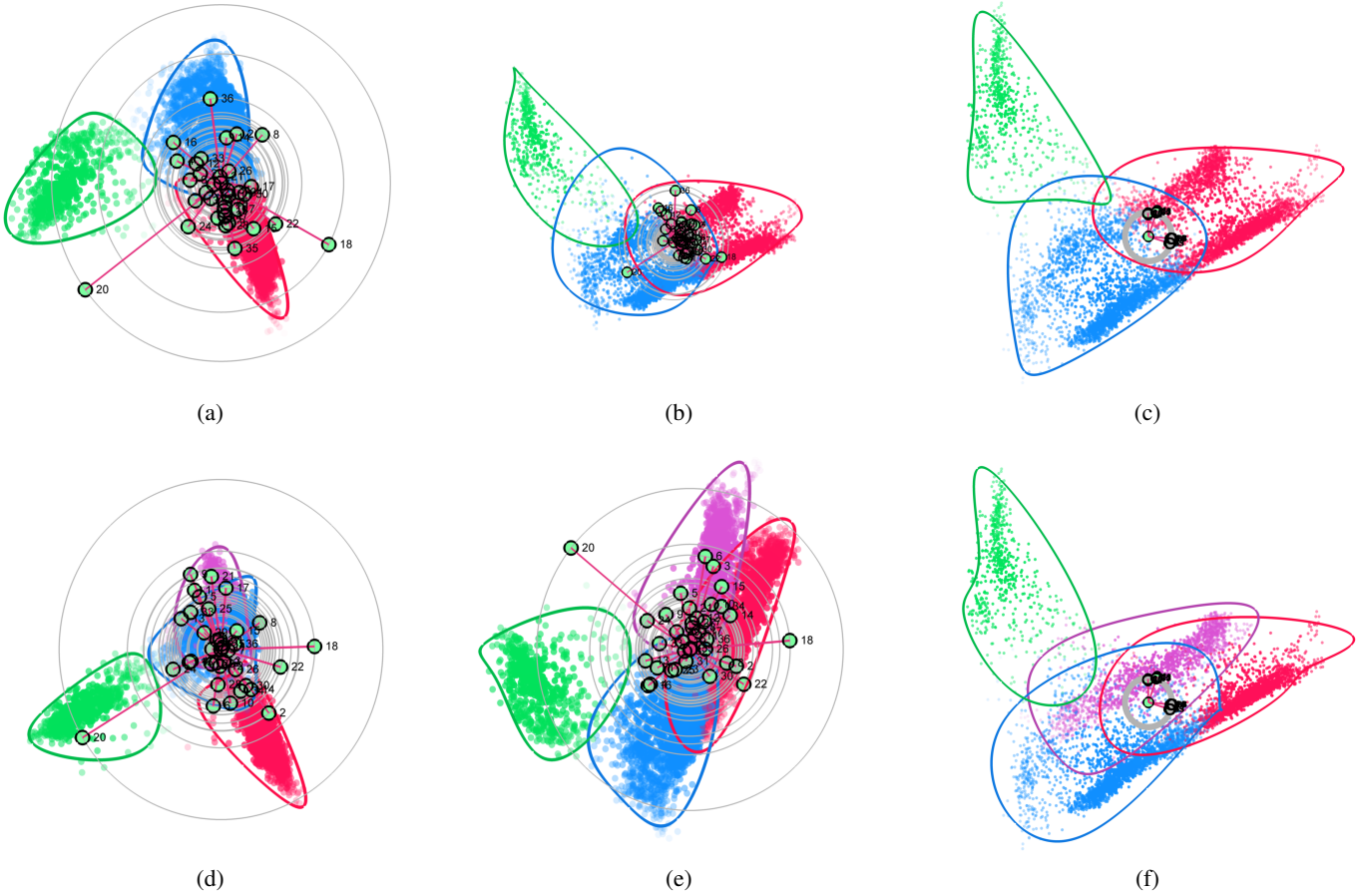


Fig. 5. Exploration of the *satimage* data set with two different numbers of clusters (3 and 4). (a,c) The DSC and PCSC views of three clusters, respectively. (b)The mid-point interpolation when (a) is morphed into (c). (d,e)The DSC views of four clusters, configured by (e_1, e_2) and (e_1, e_3), respectively. (f)The PCSC view of four clusters.

,

and the eigenvectors (e_1, e_3). In Figure 6(b), shortening the most discriminative dimension (the 20^{th} dimension) makes the green cluster mixed with the yellow cluster together while the changes to the red and magenta classes are relatively small. This indicates that the 20^{th} dimension plays an important role in discriminating green and blue clusters. Elongating the axis of the 25^{th} dimension, which is very short in Figure 6(a), results in large changes to the shapes of all five classes; see Figure 6(c). It also makes the green cluster overlap with the remaining four classes. This indicates that the 25^{th} dimension relates to all classes and has a poor discriminative ability. Figure 6(d) shows the scaling and rotation on the axis for the 18^{th} dimension, which leads to the green class highly overlapping with the red and magenta classes. Again, this suggests that the 18^{th} dimension is important in distinguishing the green class.

4.3 Linked Parallel Coordinates

Like all dimension reduction methods, star coordinates lose the original dimensional information, and thus the user cannot get the concept of data distribution of each class in the original dimensions. Thus, a linked parallel coordinates view is provided to reveal relationships between the data distribution in the original dimensions and classes of interests. Since the axis length represents the discriminative ability (DSC) or significance (PCSC) of the corresponding dimension, we allow the user to filter some dimensions with an axis length threshold and show the remaining dimensions in a linked parallel coordinates view. The user can also use a lasso to select some potential outliers in DSC and PCSC views and explore their distribution of some dimensions with parallel coordinates.

Figure 7 shows an example where a linked parallel coordinates view is used to show several explored dimensions in Figure 6. Here, the 2^{th}, 18^{th}, 20^{th} and 24^{th} are selected by filtering the axis length of Figure 6(a) with a threshold of 0.3 (mean axis length). The 25^{th} dimension is additionally put together to verify the conclusion drawn from Figure 6(c). We can observe that the green cluster is well separated in all dimensions while the other four classes have different degree of overlap in the first four dimensions. This observation is consistent with the spatial relationship between five classes shown in Figure 6(a). Unlike the first four dimensions, the data in the 25^{th} dimension is divided into three subsets, where the first and last subsets correspond to the green and blue classes, and the middle one contributes to the red, magenta and yellow classes.

5 Application and Evaluation

We have implemented and tested our prototype visualization system on a PC with an Intel Xeon E5540 2.53 GHz CPU and 4.0 GB RAM using Matlab. Our system can achieve an interactive visualization of the data sets shown in Table 1. Since both labeled and unlabeled data can be visualized with our system, we demonstrate its effectiveness in two views: LDA-supported labeled data visualization and ULDA-supported unlabeled data visualization.

5.1 DNA Data

First, we present a case-study on the splice-junction the *DNA* sequence data set from the Statlog collection [26]. The data set contains

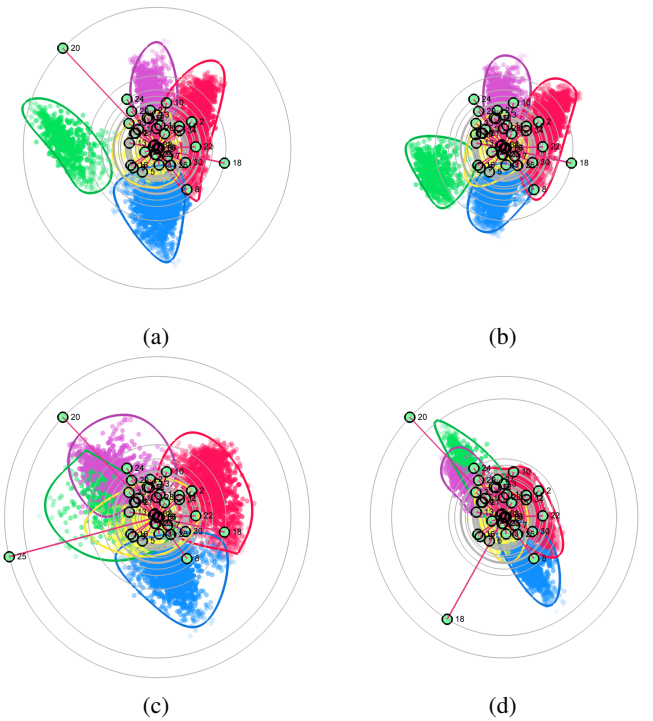


Fig. 6. Manipulating the axes of the DSC shown in (a). (a) The DSC view of the *satimage* data set (five clusters) configured by (e_1, e_3). (b) Shortening the axis of the 20^{th} dimension. (c) Enlarging the axis of the 25^{th} dimension. (d) Rotating and enlarging the axis of the 18^{th} dimension. All axis manipulations increase the overlapping among five clusters, while the relationship between these manipulated axes with class structures are different.

Table 1. Description of data sets

Data Sets	# Record	# Dim	Data Sets	# Record	# Dim
three-Gaussian	6348	3	Iris	150	3
four-Gaussian	200	4	Wine	178	13
Digits	7494	16	DNA	3186	180
Bank marketing	4521	17			

3186 DNA records and each record consists of 180 binary attributes where every 3 binary variables represent one nucleotide (A,G,T,C). According to the types of splice junctions in DNA sequence, this data has been categorized into three classes: exon/intron (EI), intron/exons (IE), and Neither. In this case study, we want to see the structure differences between classes and how each class is related to different attributes.

We started from the LDA-supported DSC view with three distinct colors (see Figure 8(a)). In DSC view, we can see three adjacent classes, where Neither (green) partially overlaps with the IE (blue) class. The Neither class appears to have a large variance, while the clusters of the EI (red) and IE (blue) classes are more coherent, although they both have some outliers shown in other two classes. The large variance implies that splice junctions of a variety of DNAs are neither EI or IE. We delved deeper into this hypothesis by examining the PCSC view shown in Figure 8(b). We found it consistent with our hypothesis, where the Neither (green) class has a large shape and highly overlaps with EI (red) and IE (blue) classes. For comparison, we also show the star coordinates with standard configuration in Figure 8(c).

Besides revealing class structures, DSC and PCSC views also indicate how each class relates with attributes. From the DSC view we can observe that the 90^{th} and 85^{th} dimensions have the largest discriminative abilities, while the 105^{th} and 93^{th} dimensions are

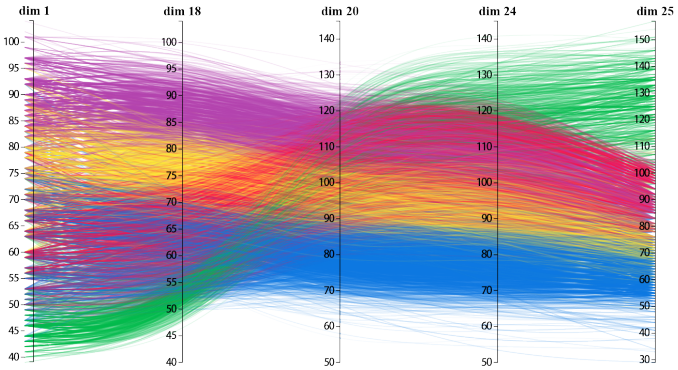


Fig. 7. Five dimensions selected from the DSC view shown in Figure 6(a) are visualized with parallel coordinates.

the most significant attributes in the PCSC view. The axes whose lengths are larger than the double mean axis length in two views are labeled in the zoomed views of Figure 8(a,b). It is interesting to find that the 85^{th}, 90^{th}, 93^th, 100^{th}, and 105^{th} dimensions appear in both views. After selecting these 5 attributes, the linked parallel coordinate views show the distribution of the data per axis of each class, see Figure 8(d,e,f). By comparing the data distributions of these three classes, we concluded that the 93^{th} dimension is the most discriminative attribute in distinguishing the EI (red) class, and the 85^{th} and 90^{th} dimensions have a similar ability in discriminating IE (blue) class, whereas the distributions of all axes in the Neither (green) class look regular. As the most significant dimension in the PCSC view, the 105^{th} dimension contributes fairly to all three classes and thus does not have a strong discriminative ability like the other four dimensions.

5.2 Bank Marketing Data

To demonstrate the effectiveness of our system in exploring unlabeled data, we conducted another user study on the *bank marketing* data set. This data set was used to predict whether a client will subscribe to a term deposit based on 16 continuous (e.g., age) and categorical (e.g., marital status) attributes in a direct marketing campaign of a Portuguese bank [28]. Here, our goal is to detect client patterns and to find out which attributes can better classify clients. Taking the client depositing behavior as another attribute, this data contains 4521 client records and 17 dimensions.

Since the number of clusters k is unknown, we tried different ks and selected the best one by comparing cluster structures in DSC views. Figure 9(a,b,c) shows the three DSC views generated by using $k = (3,4,5)$. Comparing these three views, we can see that the clusters in Figure 9(a,b) both are well separated, while the green and red clusters in Figure 9(c) appear to highly overlap each other. To verify this hypothesis, we examined the DSCM views (Figure 9(d)) of 5 clusters and found that the eigenvectors e_2, e_3 and e_4 further exacerbate the cluster overlapping where all 5 clusters are mixed together in their configured DSC views. Thus, we concluded that using 5 clusters to analyze the data is not appropriate.

The DSC view shown in Figure 9(b) clearly reveals 4 clusters and thus we do not need to examine its DSCM view. With these 4 clusters, we performed the client pattern analysis by first examining the PCSC views of the data (left in Figure 9(e)) and four clusters (right in Figure 9(e)). We can observe that four classes are separated into two groups, and each group has two adjacent classes. This is consistent with the spatial relationship between the clusters, as shown in Figure 9(b). However, each class is not compact in its PCSC view and seems to have a large variance. To investigate why these clusters have a large size, we select the top 3 significant dimensions from each PCSC view shown in Figure 9(e). Interestingly, the top 3 significant dimensions in the PCSC view of the whole data are also the most discriminative 3 dimensions shown in Figure 9(b). After removing the

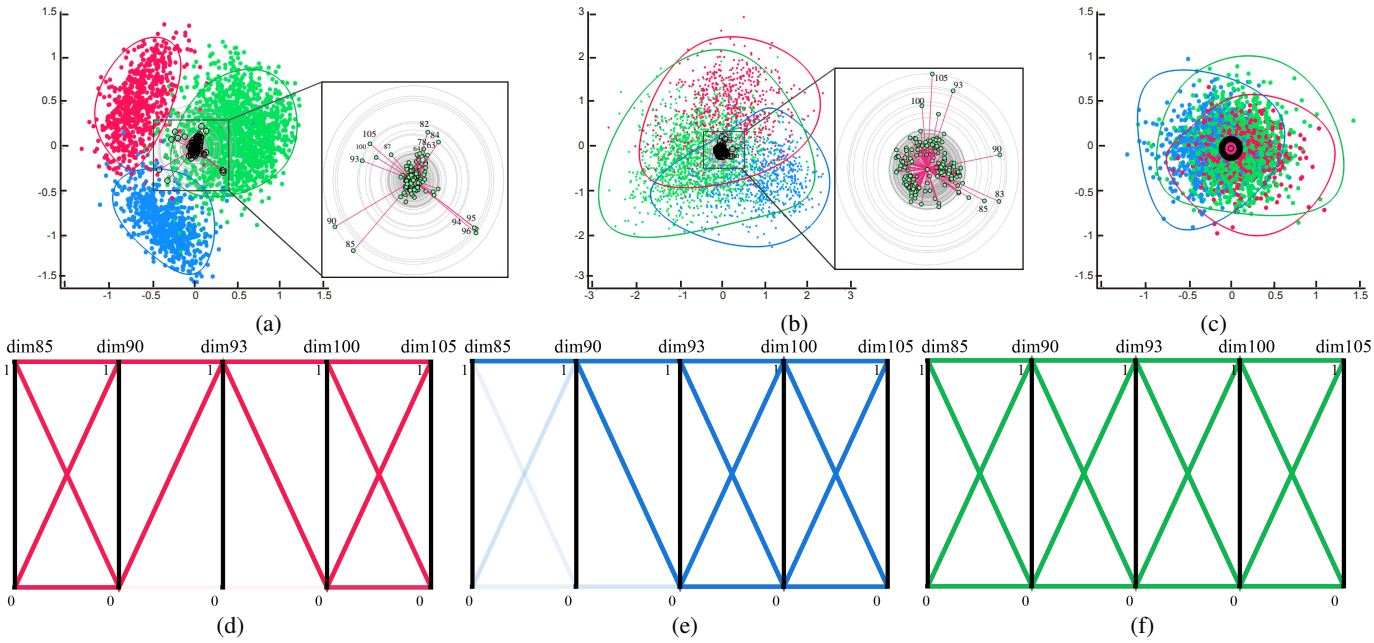


Fig. 8. Experiments on the *DNA* data set: (a)The DSC view of three clusters, where the axes are zoomed in to clearly show the axes lengths. (b)The PCSC view of three clusters, which has been zoomed out two times to display all clusters. Like (a), the axes are also zoomed in for showing the detail. (c) The star coordinates with standard configuration. (d,e,f)The linked parallel coordinates show the distributions of three classes with selected five dimensions.

,

duplicate dimensions, 7 dimensions are selected, which correspond to 7 attributes: job, martial, education, housing, loan, contact communication type, and depositing. Since all these 7 attributes are categorical variables, we concluded that the continuous attributes, such as age and averaged yearly balance, cannot characterize different classes of clients in this data set. This discovery is a little different from traditional categorization methods where age and average yearly balance are often used to classify clients [24].

With these selected 7 dimensions, Figure 9(f) shows the parallel coordinates plots of four classes. We can clearly see that the records can be separated into two groups: red and green classes, blue and magenta classes by the binary variable of housing. In each group, two classes can be distinguished by whether using the telephone communication type or not. This is an unexpected discovery, which was not reported by the state-of-the-art machine learning algorithms that analyzed the same data set [28]. A further verification of this discovery with domain scientists is needed. On the other hand, we can see that all clients in the blue class are not divorced and most of the clients in the red class are not single, while the martial status cannot be used to differentiate green and magenta classes. Except for the three attributes of housing, contact way and martial status, all categories of the remaining attributes contribute to each class and we assume this is why there is a large variation within each class.

6 Discussion

Showing the projection of star coordinates is a special case of linear dimension reduction. Our LDA-configured DSC can clearly separate different clusters of the data in 2D space. Since K-means and LDA have the same optimization objective, i.e, minimizing the within-class scatter matrix and maximizing the between-class scatter matrix [11], they are coherently combined into ULDA, which facilitates our DSC to reveal the cluster structures in unlabeled data. As shown in Figure 2, ULDA is numerically stable and can achieve convergence within a few iterations.

Compared to previous interactive subspace exploration methods [33, 40], LDA or ULDA-supported DSC provides a good starting point for cluster exploration. Combing DSC and PCSC can give the user a complete view of how the dimensions relate with the cluster structures. Note that the difference between our two-stage exploration scheme and the two-stage dimension reduction (LDA+PCA) method [7] is that the latter one performs PCA on the LDA projection rather than on the original data and thus cannot provide the complete structure of classes to the user.

Finally, we would like to mention that our DSC can also be used for initializing orthographic star coordinates (OSC) [23]. However, we do not adopt the orthography-preserving interaction in our exploration scheme, because non-manipulated axes have to be adjusted to preserve the orthography. This will hinder the user to understand how the manipulated axis affects the corresponding class structures.

7 Conclusion and Future Work

In this paper, we propose the discriminative star coordinates which can clearly characterize different clusters in 2D space. We extend our concepts to a star coordinates matrix in order to simultaneously reveal more than 3 clusters implied in the data. To facilitate the cluster structure exploration, a two-stage exploration scheme is presented, where DSC and PCSC are combined together to help the user explore class structures and the relationship between dimensions and clusters. Equipped with a set of structure-aware interactions, this exploration scheme has proven to be effective in high-dimensional data exploration.

Our approach has still some limitations, which we want to address in the future. First, the current Matlab implementation cannot support the exploration of very large data like millions of data records. Porting the LDA and ULDA computation and point rendering onto GPU is an ongoing work. Second, linear dimension reduction methods cannot handle some complex data [16] whose low-dimension projection cannot be explained by linear methods. Incorporating kernel methods to DSC is part of future work. Finally, we plan to test the effectiveness of 3D discriminative star coordinates and perform a formal user study to validate our approach in the future.

REFERENCES

[1] K. D. A., W. Müller, and H. Schumann. Visual data mining. In *Star-Report, Eurographics2002*, 2002.

[2] R. A. Becker and W. S. Cleveland. Brushing scatterplots. *Technometrics*, 29(2):127–142, 1987.

[3] A. L. Bordignon, R. Castro, H. Lopes, T. Lewiner, and G. Tavares. Exploratory visualization based on multidimensional transfer functions and star coordinates. In *Proceedings of the Brazilian Symposium on Computer Graphics and Image Processing*, pages 273–280, 2006.

[4] I. Borg and P. J. Groenen. *Modern multidimensional scaling: Theory and applications*. Springer, 2005.

[5] O. Chapelle, B. Schölkopf, A. Zien, et al. *Semi-supervised learning*. MIT press Cambridge, 2006.

[6] K. Chen and L. Liu. ivibrate: Interactive visualization-based framework for clustering large datasets. *ACM Transactions on Information Systems*, 24(2):245–294, 2006.

[7] J. Choo, S. Bohn, and H. Park. Two-stage framework for visualization of clustered high dimensional data. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, pages 67–74, 2009.

[8] J. Choo, H. Lee, J. Kihm, and H. Park. ivisclassifier: An interactive visual analytics system for classification based on supervised dimension reduction. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, pages 27–34, 2010.

[9] N. D. Cooprider and R. P. Burton. Extension of star coordinates into three dimensions. In *Proceedings of the SPIE Visualization and Data Analysis*, pages 64950–64960, 2007.

[10] I. S. Dhillon, D. S. Modha, and W. S. Spangler. Class visualization of high-dimensional data with applications. *Computational Statistics & Data Analysis*, 41(1):59–90, 2002.

[11] C. Ding and T. Li. Adaptive dimension reduction using discriminant analysis and k-means clustering. In *Proceedings of the International Conference on Machine Learning*, pages 521–528, 2007.

[12] K. Fukunaga. *Introduction to statistical pattern recognition*. Academic press, 1990.

[13] A. Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1(2):69–91, 1985.

[14] S. Jahirabadkar and P. Kulkarni. Scaf–an effective approach to classify subspace clustering algorithms. *International Journal of Data Mining & Knowledge Management Process*, 3(2), 2013.

[15] J. Johansson, P. Ljung, M. Jern, and M. Cooper. Revealing structure within clustered parallel coordinates displays. In *Proceedings of the IEEE Information Visualization Symposium*, pages 125–132, 2005.

[16] R. A. Johnson and D. W. Wichern. *Applied multivariate statistical analysis*. Prentice hall Upper Saddle River, NJ, 2002.

[17] I. Jolliffe. *Principal component analysis*. Wiley Online Library, 2005.

[18] E. Kandogan. Star coordinates: A multi-dimensional visualization technique with uniform treatment of dimensions. In *Proceedings of the IEEE Information Visualization Symposium*, volume 650, pages 9–12, 2000.

[19] E. Kandogan. Visualizing multi-dimensional clusters, trends, and outliers using star coordinates. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 107–116, 2001.

[20] H.-P. Kriegel, P. Kroger, M. Renz, and S. Wurst. A generic framework for efficient subspace clustering of high-dimensional data. In *IEEE International Conference on Data Mining,*, pages 8–16, 2005.

[21] J. B. Kruskal and M. Wish. *Multidimensional scaling*, volume 11. Sage, 1978.

[22] D. J. Lehmann, G. Albuquerque, M. Eisemann, M. Magnor, and H. Theisel. Selecting coherent and relevant plots in large scatterplot matrices. *Computer Graphics Forum*, 31(6):1895–1908, 2012.

[23] D. J. Lehmann and H. Theisel. Orthographic star coordinates. *IEEE Trans. Vis. & Comp. Graphics*, 19(12):2615–2624, 2013.

[24] B. R. Lewis and S. Spyrakopoulos. Service failures and recovery in retail banking: the customers perspective. *International Journal of Bank Marketing*, 19(1):37–48, 2001.

[25] J. MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, pages 281–297, 1967.

[26] D. Michie, D. J. Spiegelhalter, and C. C. Taylor. *Machine learning, neural and statistical classification*. Englewood Cliffs: Prentice Hall., 1994.

[27] V. Molchanov, A. Fofonov, and L. Linsen. Continuous representation of projected attribute spaces of multifields over any spatial sampling. *Computer Graphics Forum*, 32(3pt3):301–310, 2013.

[28] S. Moro, R. Laureano, and P. Cortez. Using data mining for bank direct marketing: An application of the crisp-dm methodology. In *Proceedings of the European Simulation and Modelling Conference*, pages 117–121, 2011.

[29] P. Oesterling, G. Scheuermann, S. Teresniak, G. Heyer, S. Koch, T. Ertl, and G. H. Weber. Two-stage framework for a topology-based projection and visualization of classified document collections. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, pages 91–98, 2010.

[30] L. Parsons, E. Haque, and H. Liu. Subspace clustering for high dimensional data: a review. *ACM SIGKDD Explorations Newsletter*, 6(1):90–105, 2004.

[31] J. Seo and B. Shneiderman. Interactively exploring hierarchical clustering results [gene identification]. *Computer*, 35(7):80–86, 2002.

[32] J. S. Shaik and M. Yeasin. Visualization of high dimensional data using an automated 3d star co-ordinate system. In *Neural Networks, 2006. IJCNN'06. International Joint Conference on*, pages 1339–1346, 2006.

[33] A. Tatu, F. Maas, I. Farber, E. Bertini, T. Schreck, T. Seidl, and D. Keim. Subspace search and visualization to make sense of alternative clusterings in high-dimensional data. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, pages 63–72, 2012.

[34] S. T. Teoh and K.-L. Ma. Starclass: Interactive visual classification using star coordinates. In *Proceedings of the Third SIAM International Conference on Data Mining*, 2003.

[35] R. Vidal. A tutorial on subspace clustering. *IEEE Signal Processing Magazine*, 28(2):52–68, 2010.

[36] H. Wang, S. Yan, D. Xu, X. Tang, and T. Huang. Trace ratio vs. ratio trace for dimensionality reduction. In *Proc. IEEE Conf. on Comp. Vis. and Pat. Rec.*, pages 1–8, 2007.

[37] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin. Graph embedding and extensions: a general framework for dimensionality reduction. *IEEE Trans. Pat. Ana. & Mach. Int.*, 29(1):40–51, 2007.

[38] J. Yang, A. Patro, H. Shiping, N. Mehta, M. O. Ward, and E. A. Rundensteiner. Value and relation display for interactive exploration of high dimensional datasets. In *Proceedings of the IEEE Information Visualization Symposium*, pages 73–80, 2004.

[39] L. Yu and H. Liu. Efficient feature selection via analysis of relevance and redundancy. *The Journal of Machine Learning Research*, 5:1205–1224, 2004.

[40] X. Yuan, D. Ren, Z. Wang, and C. Guo. Dimension projection matrix/tree: Interactive subspace visual exploration and analysis of high dimensional data. *IEEE Trans. Vis. & Comp. Graphics*, 19(12):2625–2633, 2013.

[41] H. Zhou, X. Yuan, H. Qu, W. Cui, and B. Chen. Visual clustering in parallel coordinates. *Computer Graphics Forum*, 27(3):1047–1054, 2008.

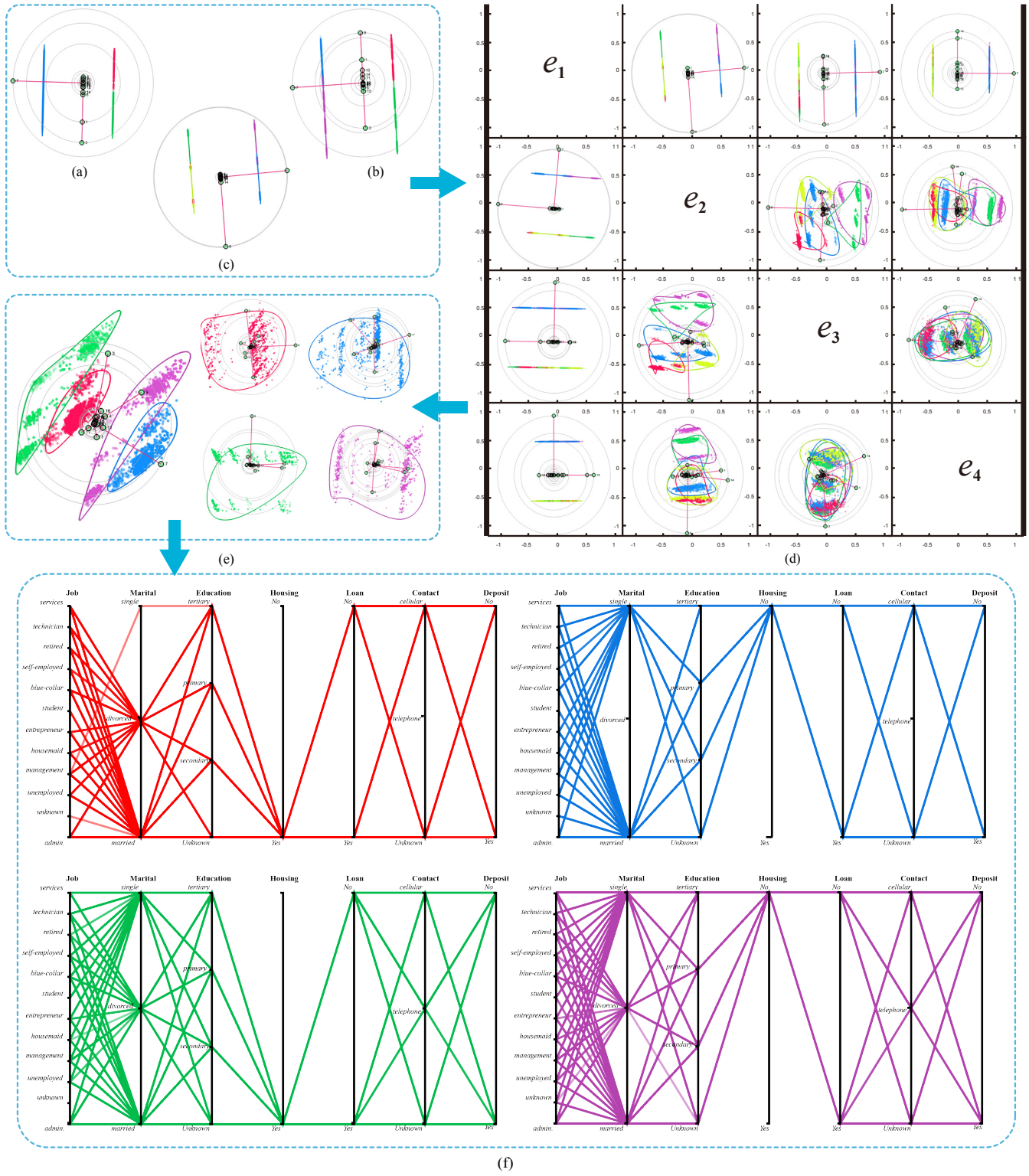


Fig. 9. Experiments on the *bank marketing* data set. (a,b,c) shows the DSC views with k 3,4,5, respectively, where the clusters in (a,b) both are well separated. (d) The DSCM view of the 5 clusters where the strong overlapping between classes makes us conclude that 5 clusters are not appropriate for client pattern analysis. (e) The 5 PCSC views (data and four classes) of the DSC view in (b). (f) The parallel coordinates views of four classes with the selected 7 dimensions.